Sticker, M. (2020). Kant, Eudaimonism, Act-Consequentialism and the Fact of Reason. *Archiv für Geschichte der Philosophie*, *102*(2), 209-241. https://doi.org/10.1515/agph-2017-0033

Kant, Eudaimonism, Act-Consequentialism and the Fact of Reason[1]

Kant considers eudaimonism as his main opponent and he assumes that his ethics is the only viable alternative to eudaimonism. He does not explicitly address theories differing from both eudaimonism and from his own. I argue that whilst Kant and Act-Consequentialists advocate different normative principles, their positions share the important abstract feature that they establish what is to be done from a rational principle and not based on what is in the self-interest of the respective agent, as Kant thinks eudaimonism does. Act-Consequentialism is thus closer to Kant's ethics than is often assumed. I will demonstrate and vindicate this point with a new interpretation of the Fact of Reason. This reading also establishes that the notion of a Fact of Reason is less contentious than many of Kant's critics believe. We should not expect that the Fact establishes Kantianism. Instead, the Fact is only supposed to count against a specific competing view of morality, namely, eudaimonism. Act-Consequentialists can accept the Fact as well.

In his provocatively titled paper "Could Kant Have Been a Utilitarian?", Richard Hare argues that Kant could and should have been a Utilitarian, but in fact he was not. Hare (1993, 8) blames this on "inbreed rigorism", which leads Kant "into bad arguments which his theory will not really support". Whilst Hare draws on the universalization requirement to show that Kant should have derived consequentialist normative principles from his Categorical Imperative, David Cummiskey (1990 and 1996) focuses particularly on Kant's Formula of Humanity and the equal value of rational nature to show that

[1] I wish to thank Joe Saunders, Matthias Hoesch, Stefano Lo Re, Pärttyli Rinne, Sarah Broadie, Leslie Stevenson, Janis Sharp, Sascha Settegast, Kiyoshi Chiba, Bruno Haas, Kerrin Jacobs, Christian Ochsner, Jens Timmermann, James Camien McGuiggan, Holmer Steinfath, Stefan Klingner, Bernd Ludwig, Phillip-Alexander Hirsch, Martin Brecher, Jacob Lautrup Kristensen, James Levine, Pauline Kleingeld, Lucas Thorpe, Courtney Fugate, Alix Cohen, Anita Leierfall and two anonymous *Archiv* referees for discussion, criticism and feedback. Earlier versions of this paper were presented at the St Andrews Kant Reading Group organized by Stefano Lo Re, at the 2016 OSZW conference in Groningen, at a workshop and a conference at the American University of Beirut, both organized by Courtney Fugate, the 2016 St Andrews Kant Reading Party organized by Lucas Sierra Velez, and at colloquia at the University of Göttingen, Waseda University, Tokyo, and Trinity College Dublin. I wish to thank everyone who attended these events and discussed my work with me. Work on this paper was facilitated by a research grand from the Irish Research Council and by the departments of philosophy of the University of Göttingen and Trinity College Dublin.

"consequentialism should follow" (Cummiskey 1996, 101) from the most abstract Kantian principles. Obviously, these views are far from what most Kantians hold about Kant's ethics. Jens Timmermann (2005, 251), for instance, argues against Hare that Kant's ethics is "utterly alien to utilitarianism and any other brand of consequentialism". Recently this debate has gained new traction in moral philosophy due to Derek Parfit's (2011) elaborate attempt to show that the most plausible forms of Kantianism and Consequentialism (as well as Kantian Contractualism) converge in a so-called "Triple Theory".[2] Parfit is concerned that if Kant and Consequentialism cannot be reconciled then we could only speak about right and wrong relative to specific ethical theories. There could thus be deep moral disagreement of a kind that makes any kind of moral objectivism impossible. For Parfit the issue of whether Kantianism and Consequentialism can be reconciled thus greatly impacts our understanding of ethical theorizing and moral truth.

My paper is an exegetical contribution to the debate on the relation between Kant's ethics and Consequentialism.[3] The underlying exegetical assumption of my investigation is that if we want to fully understand a philosopher who wrote in a historical and intellectual context different from our own, we need to understand what position(s) she considers to be her main opponent(s). Looking at a philosopher in this way often reveals important differences between the philosopher's own projects and the way her thoughts are appropriated for contemporary debates. Such an approach can afford a historically more sensitive look at this philosopher as well as provide new impulses to contemporary debates and challenge received ideas.

In my paper, I will clarify the dialectical situation in which Kant found himself, work out what position he considered his main opponent, and explain how this impacts our understanding of Kant's ethics as well as the relationship between Kant's ethics and Consequentialism. In the first section, I show that Kant sees various forms of

[2] Parfit's attempt was largely met with skepticism from Kantians. See Morgan (2009), as well as Sticker (2016a), Hoesch, Sticker (2017).

[3] I should note that I will not be using the term "deontology" in my paper, even though Kantian ethics is often subsumed under this label. Timmermann (2015, sec.IX) has recently argued persuasively that this label is inaccurate for Kant and that it is in general too broad and even misleading. I will, however, use the term "Consequentialism", even though this term itself was not used by historical advocates of this view and was originally coined by Anscombe (1958) to criticize certain forms of consequence-based reasoning.

eudaimonism as his main opponent. In the second section, I argue that impartial and maximizing Act-Consequentialism is not the kind of theory Kant explicitly engages with, and that he was even largely unaware of this position. Furthermore, I will show that on a fundamental level Kantianism and Consequentialism share some important similarities when compared to eudaimonism. In the third section, I turn to the Fact of Reason, which represents the insight that I can act on a non-self-love based rational principle and against my inclinations and even against my own long-term happiness. The Fact casts serious doubt on the conception of agency that eudaimonism draws on, but it does not establish that the only rational and non-self-love based principle is the Categorical Imperative. The Fact leaves open how to best specify this rational principle and in the fourth section I argue that Act-Consequentialism is compatible with the Fact of Reason. Finally, I show that my reading puts Kant in a good position to respond to criticism based on the false assumption that the Fact must carry all or most of the weight in Kant's argument for his ethical theory. Kantians thus should welcome my reading of the Fact, since it shows that Kant's conception of a Fact of Reason is not as contentious as some of his critics make it out to be.

Before we begin, let me make two remarks. Firstly, whilst my paper is largely exegetical, my goal is ultimately to better understand the main differences and similarities between two major ethical theories. I will therefore also motivate and support my exegesis with systematic considerations. Furthermore, there is, I believe, a mismatch between what Kant claims his prime example for the Fact of Reason shows and what it can actually show. This is so because Kant does not sufficiently consider that there are rational, non-self-love based alternatives to his own ethics. Part of my task is to rationally reconstruct Kant's notion of the Fact of Reason in the light of what his prime example, The Second Critiques Gallows Case, is indeed able to establish. I am skeptical of some of the claims Kant makes about the role of the Fact, but my skepticism ultimately serves the purpose of presenting a viable conception of the starting point of the Second Critique.

Secondly, even though I show that there is an important and fundamental feature that Kant's ethics and certain forms of Consequentialism share and that distinguishes both from eudaimonism, I do not claim that Kantianism and Consequentialism can be fully reconciled with each other. On my account, Kantianism and Consequentialism still differ in the ethical norms they prescribe as well as in a number of fundamental features, such

as the Kantian doctrine that moral worth is a question of motivation not of consequences and that for Kant the authority of moral obligations is grounded a priori and not conditional on specific circumstances and calculation of outcomes. Nonetheless, my paper fills an important gap in the literature. Those who push Kantian Consequentialism emphasize that Kant's abstract moral principle "actually provides support for a form of normative consequentialism" (Cummiskey 1990, 588).[4] By contrast, critics of Kantian Consequentialism emphasize that even if this were so, it is not the case that if two ethical theories issue similar commands concerning concrete cases then these theories must be similar in their foundational and fundamental commitments (see Timmermann 2005). It is therefore very important for this debate to look at the foundational and fundamental differences and similarities between Kantianism and Consequentialism. In the current paper, I focus on one of those: The Fact of Reason.

I. Eudaimonism

It is a popular and widely accepted notion that Kant's main opponent is Consequentialism, an ethical theory that commands that we maximally and impartially promote the good or goods. After all, contemporary Kantians are particularly keen on presenting arguments against Consequentialist notions such as that the rights of individuals can be sacrificed in the pursuit of a greater good for the majority. This focus is reflected in and reinforced by the way Kantianism and Consequentialism are often treated as the major opposing poles in current ethics courses and in ethical debates.[5] The current opposition between Kantians and Consequentialists, however, has led to a misunderstanding of the dialectical situation Kant found himself in. When we look at Kant's engagement with other thinkers it becomes apparent that he directs almost all of his critical attention to forms of *eudaimonism*. *Eudaimonism* is the view that it is rational for agents to do what morality commands because this is in some form conducive to the agent's *own* well-being. Conforming to the commands of morality would thus be a smart

[4] Forschler (2013) and Cummiskey (1990, sec.2) explicitly distinguish between a Consequentialism on the level of norms and on the level of foundation of these norms. According to Cummiskey, Kant is a consequentialist in the former sense only.

5 An important figure here is certainly Rawls (1971) who advocated a prominent position in a Kantian tradition and saw Utilitarians as his main opponents.

way of pursuing personal happiness.

In a lecture from 1784, Kant laments that virtue has never been presented in its full purity by philosophers (XXIX:626.10-2). In a *Groundwork* footnote, addressed to the popular philosopher and educator Johann Georg Sulzer, Kant argues that this fundamental failure is the reason why moral philosophy, so far, has failed to morally improve anyone (IV:411.fn, see also IV:426.12-21, 432.25-433.11, VIII:287.22-288.29). Kant believes that the fundamental mistake of his philosophical predecessors and contemporaries is that when they present duty they are guilty of "peddling the empirical mixed with the rational" (IV:388.26-7) or of "dally[ing] with pathological impulses" (V:85.26-7). They believe that external incentives that draw on an agent's desires to avoid punishment and to obtain reward make it more likely that agents conform to duty. Kant warns that this, in fact, deprives actions of all moral worth. He even claims that a moral theory founded on anything impure is not merely mistaken but does not even qualify as "moral philosophy" since such a position undermines the goal of moral philosophy, namely, to make people better (IV:390.13-8).[6]

Kant sees eudaimonism in the form of popular philosophy, moral sense theory, (quasi-)religious doctrine, perfectionism and practical philosophy inspired by ancient virtue ethics as his main opponent. He emphasizes that moral philosophy is never a "*doctrine of happiness*" or an "instruction in how to become happy" as eudaimonism has it (V:130.17-8). In *Groundwork* II, Kant's main antagonists are popular philosophers, contemporaries of his who advocate forms of eudaimonism. In the Second Critique, his antagonists are chiefly historical figures, such as the ancient Greek and Roman virtue ethicists and their various eudaemonist conceptions as well as Hume's empiricism and corresponding theory of action that leaves no room for actions from pure practical reason (V:13.6-21, 50.32-57.13, 93.30-106.9).[7] All of these theories, according to Kant, have in common that they hold that moral action leads to an agent's personal happiness – either

---

[6] See also IV:411.14-6, 442.6-22, V:116.14-117.24, 155.2-11, 157.7-21, VI:376.34-377.12, 482.30-483.8, VIII:395-6fn. In IV:390.19-391.15,

7 In IV:441.29-444.27, V:40.-41.38, XXVII:252.6–255.28, XXIX:621–9, Kant systematizes and discusses the different impure and heteronomous principles of uncritical moral philosophies. According to Kant, all of these principles lack the crucial insight that rational agents can be subject to a universal and yet self-imposed law and thus can act from something else than external "stimulation or constraint" (IV:433.1) or "alien impulse" (IV:444.26).

immediately, or in the long run, or in an afterlife[8] – and that this is where the authority of morality derives from.[9]

Kant's readiness to class everyone else as a eudaimonist becomes particularly apparent in a *Groundwork* footnote in which he subsumes moral sense and compassion under the principle of one's own happiness "because every empirical interest promises to contribute to our well-being by the agreeableness that something has to offer" (IV:442fn.).[10] Kant here overlooks that moral sense and certainly compassion can, as a matter of fact, incentivize actions that greatly go against one's own interests. It is unlikely that in all of these cases there is always the promise of a larger contribution to the agent's own well-being in the background. Furthermore, some remarks Kant makes about lying promises we can best understand in the context of his criticism of eudaimonism: he emphasizes, for instance, that "it is most uncertain" (VIII:287.3) that one would actually be furthering one's own happiness by keeping a deposit instead of returning it to its rightful owner (see also IV:402.36-403.2). This is presumably an attempt to criticize eudaimonism internally. Due to the unpredictability of consequences (see my sec.II) an ethical theory that prescribes to do what is foreseeably good for me might occasionally lead me to act in ways that are bad for me and it would have thus been better not to do what the theory commands. By contrast, it is a course of action far less riddled with contingencies to orient oneself on a priori grounded ethical rules instead of foreseeable consequences.

In the *Doctrine of Virtue*, Kant suggests that, at bottom, the eudaimonist's mistake is to explain actions only via "physiological explanations" or in terms of pleasure and pain. The eudaimonist lacks a conception of transcendental freedom (VI:378.19-31, see also

---

[8] Kant clearly rejects theological positions that claim that it is rational for agents to obey moral laws because of reward in an afterlife (V:146.29-148.5). As we will see below when we discuss the philosopher and theologian Crusius, it is not clear, though, that theological ethicists are committed to such a position. They are more complex than secular eudaimonists.

9 See for instance: "Now the *eudaimonist* says: this delight, this happiness is really his motive for acting virtuously. The concept of duty does not determine his will directly, he is moved to do his duty only by means of the happiness he anticipates" (VI:377.24-7).

[10] Kant repeats in the Second Critique that moral sense doctrine reduces morality "to desire for one's own happiness" (V:38.16-7). However, later in the *Doctrine of Virtue* Kant articulates a more nuanced conception of moral feeling (VI:399.19-400.20).

VIII:28512-22). Eudaimonism as a normative theory typically draws on a form of psychological hedonism or psychological eudaimonism[11], the doctrine that every human pursuit strives for one's own personal happiness. A eudaimonist sees the role of ethics as giving advice on how to best achieve personal happiness (VIII:287.22-288.29). Kant acknowledges that eudaimonists do not necessarily maintain that agents always act *directly* for the sake of personal happiness (see VI:377.24-7). Altruistic, noble or virtuous actions are possible for a eudaimonist, but if an agent wonders why it is rational to be altruistic, noble or virtuous, the answer can ultimately only be given with appeal to her own happiness. Furthermore, this ultimate answer presumably does not necessarily have to conceive of noble or virtuous actions as a mere *means* to personal happiness. Noble or virtuous actions might be a *part or actualization* of personal happiness.

This, of course, is how *Kant* thinks of eudaimonism but not necessarily how eudaimonists themselves conceive of their theories. Most importantly, they might reject the idea that eudaimonism is a form of refined egoism. After all, there are teleological theories that prescribe the promotion of eudaimonia *for everyone*. For our purpose, these theories might be best understood as forms of impartial Consequentialism with a rich conception of the good. Furthermore, many eudaimonists insist that on their conception *eudaimonia* has an ineradicable social or political element and that their conception therefore cannot be a form of egoism in the standard sense.[12] I will show in the next section that Act-Consequentialism does not neatly fit into Kant's categorizations of ethical theories, and we should be open that there can also be eudaimonist conceptions for which this is the case because they are ultimately not self-love based. However, these are not the kinds of conception Kant directly takes issue with in his writings and lectures.

Before I turn to Consequentialism, let me add two necessary qualifications: Firstly,

---

11 The term "psychological eudaimonism" is Timmermann's (2007a, 168) label for the psychological basis of the ethics of the popular philosopher Garve. See for instance: "*From happiness* in the most general sense arises the motives for every effort and so too for observance of the moral law" (Christian Garve: *Versuche über verschiedne Gegenstände aus der Moral und Literatur*, quoted from Kant VIII:281.33-5).

12 We can find such a conception for instance in the *Nicomachean Ethics*. Aristotle claims that it is the task of the statesmen to facilitate *eudaimonia* for others and that the well-being of the community is more important than that of individuals (1094b5-10). This, however, might be in tension with book X's focus on individual contemplation (1177a27-b1, 1177b33-1178a1). I am grateful to Joe Saunders, Sarah Broadie and Sascha Settegast for discussion of what eudaimonists are committed to.

Kant is aware that the *Stoics* acknowledge a stern conception of duty, even to the extent that Kant criticizes them for straining "the moral capacity of the *human being* […] far beyond all the limits of his human nature" (V:127.2-4). They are, however, eudaemonists at least in the sense that for them personal happiness and virtue are analytically connected. Personal happiness is "contained in consciousness of one's virtue" (V:112.16, see also V:11.18-28, 112.7-8, 115.26, XXIX.600.10-2). Since virtue and happiness are analytically connected, doing something for the sake of virtue is synonymous to doing it for the sake of personal happiness. This entails that the Stoics cannot account for cases in which a morally good agent is unhappy and a villain happy (IV:442.16-7, V:60.26-36, 127.7-16, XXIX:623.20-33).

Secondly, *Christianity* represents morality in an even purer and stricter form than Stoicism does (V:127-8fn.). Morality does not promise reward (V:128.16-7, VIII:339.4-19), and Christianity is "autonomy of pure practical reason by itself" (V:129.8-9). I take it that Kant here does not intend to describe and criticize an ethical theory different from his own but that he projects his own theory into Christianity and aims to show that Christian doctrine, correctly understood, agrees with his theory.[13]

The Christian thinker who comes closest to an ethics of autonomy and who Kant is familiar with is the 18th Century theologian and philosopher Christian Crusius. Crusius believes that moral principles are given by God, but he maintains explicitly that reward and punishment cannot be the reason why moral laws are binding (*Anweisung* §133, §194). For Crusius moral principles are not self-imposed but agents are nonetheless required to follow these laws for other reasons than self-love. According to the Second Critique, however, Crusius advocates material, objective, external determining grounds, just like other theological ethicists do (V:40). This is incorrect and uncharitable given that Crusius emphasizes that morality is not concerned with punishment or reward. Kant overlooks how close Crusius already is to an ethics of autonomy.[14] This might be so

[13] Elsewhere, he claims that all ethical theories based on divine will are heteronomous (IV:443.3-36). See, by contrast, Kant's more nuanced take in the Religion where he distinguishes between a "*religion of rogation*" and a "*moral religion*, i.e. the religion of *good life-conduct*" (VI:51.27-8).

14 The Crusius literature frequently stresses the similarities between Kant and Crusius (see Schmucker 1961, 81, Benden 1972, ch.7, esp.208-10, Fugate 2009, 281). See, however, also Sticker (forthcoming) who stresses and discusses differences in Kant's and Crusius' respective conceptions of conscience and moral cognition. I am grateful to Stefan Klingner for bringing Crusius' unique position to my attention.

because occasionally Crusius still does make claims that seem to be incompatible with non-eudaimonism, for instance, when he says that both prudence and duty aim at the same object: happiness (*Begriff* §36, 168, see also *Begriff* §189-91). More importantly, however, Kant's oversight of the parallels between Crusius' ethics and his own is a clear symptom of his general suspicion that everyone else is a eudaemonist (even when a careful reading would show otherwise).[15]

II. Act-Consequentialism

After having established that Kant's chief opponent is eudaimonism, we will now look at a class of ethical theories that differ both from eudaimonism and from Kant's own theory: Consequentialism. I will show that Kant does not address and is unaware of Act-Consequentialism as we understand it today as a theory that holds that "an act is good if and only if that act maximizes the good, that is, if and only if the total amount of good for all minus the total amount of bad for all is greater than this net amount for any incompatible act available to the agent on that occasion" (Sinnott-Armstrong 2015). Three things are important for us about this characterization:

Firstly, Consequentialism is about the good for *all* or about an *impartial* good[16] and it requires that we *maximize* this good.[17] This means that Consequentialism can require great sacrifices of personal goods from affluent agents. Kant famously distinguishes between theories that require actions from self-love or from empirical determining grounds and theories that require actions from duty or from pure rational determining

15 Wood (2000, 261) argues that "Kant was among the first to break decisively with the eudaimonistic tradition of classical ethics". He, however, also notes that the Cambridge Platonists and Ralph Cudworth in particular anticipated this departure (Wood 2000, fn.1).

16 Timmermann (2005, sec.1) argues that one of the major differences between Kant and Utilitarianism is that the latter theory is "strongly egalitarian" (Timmermann 2005, 246) or demands that happiness is to be promoted impartially. Kant's ethical theory, by contrast, "provides an immediate justification for some kind of partiality" (Timmermann 2005, 249-50).

17 Cummiskey (1996, 89) in his case for Kantian Consequentialism claims to have discovered a sense in which Kant too is a maximizer: His theory requires "the maximal promotion for the conditions that are necessary for the flourishing of rational agency". According to Dean (2000, 30), the issue of maximization is, alongside the question of how to understand the equal value of agents, the main point of contention between Kantian Consequentialists and standard Kantians.

grounds. The former is eudaimonism, the latter Kant's own ethics. Kant seems to think that the distinction is exhaustive. Whilst Act-Consequentialism does not share the Kantian formal conception of duty, it is, however, odd to think of it as a theory based on the agent's self-love. In fact, in the wake of Singer's (1971) famous "Famine, Affluence and Morality", Consequentialism has even come under fire for imposing excessive restrictions on agents' pursuit of personal well-being.[18] Actions prescribed by Act-Consequentialism are often not in the service of an agent's self-love, at least not for agents who are well off and who are therefore required to sacrifice much.

Secondly, Consequentialism is sometimes associated with a narrow and monolithic conception of the good, classically pleasure and absence of pain. The above characterization does remain neutral concerning how to spell out the good(s) to be promoted and I speak of "Consequentialism" rather than "Utilitarianism" to avoid commitment to any specific and narrow conception of the good. Consequentialists can, for instance, make fair distribution part of their conception of the good.

Thirdly, the most important historical *predecessors* for our modern impartial and maximizing conception of Consequentialism are Jeremy Bentham (1789) and Henry Sidgwick (1907). There is no indication that Kant was acquainted with the works of Bentham and he obviously could not have been acquainted with Sidgwick. In this sense, it is trivially true that Kant was unaware of our contemporary notion of Consequentialism. However, the main ideas behind this notion – that we can assume an impartial standpoint according to which everyone (including the agent herself) counts as much as everyone else and that we should do as much good as we can – represent natural and intuitive ways of thinking about morality and these ideas are not simply the product of philosophical theorizing.[19] It is therefore not anachronistic to ask whether and how these ideas are reflected in Kant's discussion of the good.

I shall bracket another important Consequentialist figure, John Stuart Mill, since he

---

18 In his paper, Singer does not intend to present a specifically Consequentialist theory. Yet, his theory is often considered a paradigm of Consequentialism.

19 Even prior to the late 18th/early 19th century thinkers such as Hobbes, Hume, Hutcheson, Shaftesbury, Beccaria and Helvetius incorporated Consequentialist elements into their theories, as Sidgwick (1907, 423-6) argues. Sidgwick also argues that common sense is "at least unconsciously Utilitarian" (Sidgwick 1907, 424).

is rather a Rule-Consequentialist (Mill 1861, ch.2.24-5, and ch.5), and I will not address Rule-Consequentialism. I share the worry that Rule-Consequentialism either collapses into Act-Consequentialism (if its rules are so specific that in every situation the agent is required to bring about the maximum good she can) or can be criticized from a Consequentialist perspective for making a fetish of rules (when it forbids an action that would maximally promote the good).[20] I focus on *Act*-Consequentialism, since this form of Consequentialism stayed truer to the central Consequentialist ideas that the good ought to be promoted impartially and as much as possible.

Philosophers often assume that Kant's main antagonists are Consequentialist theories, since Kant frequently argues against basing ethical assessment on the consequences of actions. Most importantly, he argues that since consequences are contingent we could never be *certain* what to do if we followed a consequence-based theory (IV:418.1-37, VI:215.24-216.6). This problem, sometimes referred to as the "Cluelessness Objection"[21], is still a serious challenge for contemporary Consequentialists.

It is telling that in Kant's own discussions of consequences it is usually left unspecified *who* is affected by these consequences or *whose* good an agent intends to promote: her own, that of specific others or everyone's?[22] Importantly, in those passages

[20] The *locus classicus* for this objection is Lyons (1965). There is a larger issue in the background here: If the Fact of Reason is, as I will argue below (sec.III-IV), supposed to discredit ethical theories that assume that agents can only ever act for the sake of their personal happiness, then potentially many other theories can accept the Fact, namely, all theories that do not subscribe to a form of psychological eudaimonism/hedonism. These could be forms of Rule-Consequentialism, Scanlonian Contractualism, Intuitionism, and many more. I do not mean to deny this. Some of these theories, such as Scanlonian Contractualism, are Kantian in nature and the claim that they could accept the Fact of Reason would be somewhat unsurprising. Other theories, for instance forms of intuitionism, Rule-Consequentialism, etc. would require substantial separate debates. In this paper, I focus on Act-Consequentialism, since it is one of the major and most influential ethical theories to date; and since this theory is frequently seen as sharply at odds with Kantianism. Showing that even Act-Consequentialists can accept the Fact of Reason is therefore the most effective way to show that the Fact does not only support Kant's own ethics or Kantian theories.

21 The term was coined by Lenman (2000) to refer to the problem that due to the difficulties in determining the consequences of our actions, Consequentialism does not offer substantial clues for answering ethical questions.

[22] For instance, in a *Metaphysics of Moral's* discussion in which Kant seeks to distinguish his own project

in which Kant is more specific, he has in mind how the consequences of an action would impact the agent's *personal* happiness. In the *Groundwork*, for instance, Kant explicitly speaks of how the indeterminacy of consequences affects "*my* present and every future condition" (IV:418.8, my emphasis).[23] Furthermore, in one of the central passages for the indeterminacy of consequences (V:36.28-37.13), Kant adds to the indeterminacy problem that a "command that everyone should seek to make himself happy would be foolish, for one never commands to someone what he unavoidably wants already" (V:37.5-7). However, it is *one's own happiness* that, according to Kant, an agent unavoidably wants, not universal happiness (V:25.12-4, VI:386.3-7, 387.26-9). This strongly suggests that Kant's discussion of the indeterminacy of happiness here is concerned with the pursuit of one's *own* happiness.

Furthermore, in two different *Groundwork* passages, Kant discusses candidates for unconditional or absolute value, but strikingly universal happiness is not among them. Firstly, in the beginning of *Groundwork* I, Kant dismisses the notion that "well-being and contentment with one's condition, under the name of *happiness*" can be unconditionally good. Kant here has in mind personal happiness, not universal happiness, as becomes apparent from his argument that happiness is not unconditionally good since "a rational impartial spectator can nevermore take any delight in the sight of the uninterrupted prosperity of a being adorned with no feature of a pure and good will" (IV:393.15-22). The spectator here does not observe and assess the state of the world as a whole; he only observes an *individual* ("a being" [eines Wesens]), namely, the agent who made the principle of personal well-being her highest imperative. This passage thus makes a point about the happiness of individuals as a candidate for the unconditional good, not about the maximum happiness of everyone. Secondly, in the *Groundwork* II argument that

---

from a doctrine of happiness, he speaks of this doctrine as concerned with both: the "means for achieving one's lasting enjoyment" and "what brings us joy" (VI:215.24-216.6, my emphasis). The former suggests that this is about the enjoyment of the agent herself, the latter that it might be about the enjoyment of everyone. It should be noted that Gregor (1996, 370) does not translate the "man" ("one's") in the former passage. Thus, her translation reads as if *both* passages are concerned with *our* enjoyment or universal enjoyment.

23 It also becomes apparent in the examples Kant provides for the indeterminacy of happiness that he sees this indeterminacy as a problem for an agent's *personal* happiness (IV:399.12-21, 402.16-403.17, 418.8-24, VIII:286.8-287.21).

humanity is of absolute value, Kant dismisses three other candidates for absolute value: Objects of inclinations, inclinations themselves and non-rational beings (IV:428.7-33). He does not entertain universal happiness as an option.

Finally, in the Second Critique, Kant presents the example of an acquaintance who tries to justify having given false testimony by appeal to the supposed duty of "his own happiness and […] all the advantages he had acquired by doing so" (V:35.19-22). Strikingly, Kant here does not discuss an acquaintance who lies for the sake of *everyone's* happiness, even though this would have been a much more important challenge to dispel for Kant, since this excuse does not sound quite so absurd as the one of the lying egoist.[24]

There is only one place in Kant's published works in which he directly discusses universal happiness, namely, in V:36.10-11. Kant here entertains that "*universal* happiness" could be incorporated into one's maxims.[25] Kant dismisses this possibility quickly: Prescriptions based on one's goal to promote universal happiness would be based on experience, they would be contingent, a posteriori and lack necessity (V:36.11-24) and ultimately only concern what is agreeable and disagreeable (V:58.10-35, 63.21-28). The way Kant introduces universal happiness ("even if *universal* happiness were made the object [of the will]" V:36.10-11) suggests that Kant does think of this as a far-fetched position ("even if") and not a position anyone actually holds. Furthermore, he does not indicate any source for this position, as if he considers it a merely logical possibility.[26]

---

[24] See also V:430.6-431.35 where Kant rejects the idea that material determining grounds can make an agent an ultimate end of creation. He does not discuss universal happiness or other goods as a candidate for material determining grounds, but only happiness as an "*idea* of a state" of ideal fit between the world and an agent's *own* instincts ("his instincts" V:430.7).

25 See also the brief mention of "universal happiness" in a *Reflection* (XIX:279.31) and Kant's brief remark in the Second Critique that ancient philosophers took an interest in the "general good" [allgemeinen Guten] (V:109.7). The latter might be in acknowledgement of the fact that Plato, Aristotle and the Stoics are not ethical egoists in a straightforward sense. In the *Religion*, Kant, on a number of occasions, uses the term "best of the world" [das Weltbeste] (VI:61.20, 32, 173.26, see also VIII:147.4). For instance, in VI:61.20, Kant claims that Jesus sacrificed his life for the *Weltbeste*. "Weltbeste" here presumably refers to forgiveness of sins, not to universal happiness in our modern, secular sense.

[26] A possible source is Pistorius, though I will argue below that Kant did not understand Pistorius' criticism as a challenge from impartial Consequentialism.

He presumably thinks the position is implausible from the start, since peoples' happiness is based on their individual faculties of desire, "which can never be assumed to be universally directed to the same object" (V:26.3-6). The desire for happiness can therefore never pass of as a "universal *practical law*".[27] Kant assumes that the fact that different people have different and incompatible notions of happiness and that different things make them happy, makes talk about universal happiness spurious.[28]

Kant is right that it is, as a matter of fact, impossible that we achieve a state of the world in which everyone is maximally happy, since some of the goods that make agents happy are limited and satisfying some of the preferences of one agent is incompatible with satisfying some of the preferences of another agent, since both agents might be competing for the same thing (V:28.17-28). Furthermore, some agents might even desire other agents' unhappiness or suffering. However, we need to distinguish between universal and *impartial* happiness. The former requires that I promote the complete happiness of every agent, a state that, due to incompatible preferences, is impossible to achieve. The latter only requires that I promote *overall* happiness and that I do so with no special consideration (or no special considerations that lack an impartial justification) for my own happiness and the happiness of loved ones. The latter is typically what Consequentialists aim for, since they accept the idea that agents cannot be obligated to do the impossible. If we aim for impartial happiness differences in what makes people happy and even incompatible preferences only imply that it is *difficult* to promote overall happiness and that we have to know much about human psychology and the specific circumstance of agents and that we have to decide impartially between incompatible preferences. There is nothing conceptually confused or spurious about agents making it their maxim to do as much as they can to impartially promote happiness.[29]

That a discussion of impartial happiness is largely absent from Kant's engagement with other ethical theories is all the more revealing, since there is one instance in which Kant

[27] V:28.7, see also IV:417.27-419.2, V:21.32-22.3, 28.23-8, 430.6-16, VI:215.24-216.6.

[28] In addition, Kant is skeptical that happiness is attainable (V:129.4-6, 430.16-23), and he thinks that it is an end that opens individuals up to contingencies and uncertainties (IV:417.27-419.2).

[29] I am grateful to an anonymous referee for pressing me to say more about how Consequentialists could deal with Kant's charges against the notion of universal happiness.

was directly confronted with an impartial form of Consequentialism. In 1786, Herman Andreas Pistorius claimed in a review of the *Groundwork* that, instead of with a good will, ethics should start with a "highest and absolute good" (p.8-9).[30] A good will then "would be that will whose maxim is: do that which is in conformity and agreement with your and simultaneously the interest of all rational beings" (p.9). Pistorius shortly after clarifies that "if apparent collisions of one's own and the common interest arise, the latter comes before the former" (p.9). Pistorius does not explicitly advocate an *Act*-Consequentialist conception (he implicitly agrees with Kant's focus on maxims or rules) and he also does not explicitly present a maximizing form of Consequentialism. Yet, Pistorius confronts Kant with a way of thinking about ethics that starts with the notion of a material good ("the highest and absolute good") that can require that you sacrifice your own well-being for the common good.

Kant takes up the challenge to clarify the place of a highest good in the Second Critique.[31] Interestingly, he presents the highest good here as a combination of virtue and "one's own happiness".[32] Even here Kant does not entertain a role for universal happiness. It seems that Kant almost willfully misreads Pistorius, which is very telling. Kant assumes that if Pistorius is not with him, then he ultimately has to base ethics on personal happiness, and thus, in response to Pistorius' challenge, he shows that there is an important, albeit subordinate, place in his ethics for *personal* happiness. Kant seems oblivious to the possibility that we can base ethics on *impartial* happiness or, more generally, on impartial promotion of goods.[33]

---

[30] I quote Pistorius according to an unpublished translation by Michael Walschots.

[31] That Kant addresses Pistorius alongside ancient eudaimonism reinforces the notion that Kant thinks Pistorius' challenge comes from a eudaemonist background, not from an impartial one.

[32] V:112.28, see also V:110.27-111.5, 119.1-23, 124.21-5, 129.33-130.16. That the highest good only encompasses an agent's personal happiness, as opposed to everyone's happiness, we can already find in the First Critiques' *Canon* (A/B 805-6/833-4, 813/841). I owe some of these references to an unpublished paper by Jens Timmermann. Kant's focus on personal happiness is also consistent with his claims that an agent is not required to "renounce claims to happiness but only that as soon as duty is in question, one should take no account of them" (V:93.11-5, see also VIII:278.15-9, 283.15-9). Kant never requires that agents adopt an attitude of indifference towards their personal well-being.

[33] In the *Common Saying*, the "*highest good* possible in the world" is explicitly presented as a response to Christian Garve and his eudaimonism (VIII:279.20). Kant here describes "universal happiness" (VIII:279.21) in conformity with morality as one of the elements of the highest good. It should be noted,

It is hardly an accident that Kant neglects impartial happiness and only addresses the personal happiness of individual agents. For Kant theories either have a formal *principium diiudicationis*, or a material one and they derive duties from goods (a matter), which precedes all moral principles. In the Second Critique, he claims that "[a]ll material principles as such are, without exception, of one and the same kind and come under the general principle of self-love or one's *own* happiness".[34] These principles can only determine the will via pleasure and pain for the agent, i.e., via an agent's self-love.[35] There is a supposedly exhaustive dichotomy between morality understood in a formal sense and one's own happiness or the wish to avoid pain/seek pleasure *for oneself* (V:35.7-18, 62.15-26). Kant thinks of how an action affects "*my* existence" (V:63.14-21, my emphasis) as the only thing other than respect for the moral law that can move to action.[36] He cannot fathom that an agent does not act from the Categorical Imperative but yet also not from mere self-interest, and his conception of material principles seems to rule this out from the start. Material principles give "nothing other than heteronomy" (IV:444.3), and Kant believes that "*all* the errors of philosophers with respect to the supreme principle of morals" (V:64.6-9, my emphasis) are due to philosophers basing their theory on a notion of material goodness. By contrast, Kant is unconcerned about philosophers picking a wrong *formal* principle; or rather he thinks that there is only one possible formal

---

however, that the German has "allgemeine, jener gemäße Glückseligkeit", which is not quite as straightforward as "universal happiness". Moran (2012, ch.1) argues that from the *Common Saying* onwards Kant understands the highest good as concerned with morality and the happiness of *everyone*. I cannot discuss here developmental aspects of the highest good, and will focus on the Second Critique conception.

[34] V:22.6-8, my emphasis, see also IV:453.25-31, V:21.17-25.10, 34.2-11, 35.7-11.

[35] In the *Common Saying*, Kant attributes the notion that ethics should start from a material good and derive duties from this good to the eudaimonist Garve who argues that if a theory does not start from a good, it cannot provide any incentives for actions (VIII:281.33-282.2).

36 DeWitt (2014, sec.2) argues persuasively that Kant cannot conceive of material goods as impartial because the faculty of feeling, which is responsible for motivation, "judges the object of a representation of cognition to be good *in relation to a subject*, and so brings the representation under the active scope of the faculty of desire, our source of causality for making the object of that representation actual" (De Witt 2013, 41-2, my emphasis). Judgments that can motivate agents always concern how an object would affect the agent who judges; either in the form of gratification or, in the case of morality, in the form of formal pleasures directed at concepts that are internal to the activity of the will (see also V:27.7-12, VIII:395-6fn.).

principle to pick.

Act-Consequentialism is commonly understood as starting from material good(s).[37] Kant emphasizes that the assumption of a substantial good that precedes the moral law "would always produce heteronomy and supplant the moral principle" (V:109.25-33). He expresses concern about material principles, because they can only determine the will through "feeling, which is always empirical" (V:64.21-2, see also V:21.17-31). All material principles "turn on the principle *of one's own happiness*" (V:34.7-11). It seems that Kant is not worried about material determining grounds as such, but rather about how they affect moral motivation. He believes that we cannot have material starting points without a form of eudaimonism, according to which agents can only be moved to actions through empirical feelings or the expectation of pleasure. This, however, seems false for material determining grounds that require that one promotes *impartial* happiness (and other goods), potentially at the expense of one's own happiness. At least some, and, I think in fact many, acts of charity motivated by the good consequences they would bring about, are more naturally understood as expressions of what the charitable agent thinks is the right thing to do or what she thinks impartiality requires of her, rather than as actions that are supposed to further one's self-interest.[38]

---

37 Sidgwick (1907, 391-2) would be a classical advocate of such a Consequentialism. However, not all Consequentialists are committed to a priority of the good. Hare (1993, esp.13-4), for instance, argues that Consequentialist principles can be based on formal rational requirements, such as universalizability. There might be two types of Consequentialists: Those who start directly from material goods and those who start from rational principles and derive material principles from this. See also Forschler (2013, 89-90) for this distinction. Hill (1999, 150-1) reminds us that Consequentialists do not have to hold that we ought to maximize happiness *because* happiness is intrinsically good. There can be formal justifications for the pursuit of impartial happiness.

38 I admit, this is an empirical claim. Obviously, I cannot prove that it is not the case that Consequentialists secretly expect reward for their moral actions and only act because of this expectation. However, if we look at prominent contemporary Consequentialists such as Peter Singer or Derek Parfit, then hope for reward plays little to no role for their theories. What carries the weight is rather the badness of suffering, and a notion of practical rationality understood as a (largely) impartial standard. Moreover, MacFarquhar (2015) recently presented a series of short biographies of the lives of morally extraordinary people who donated almost everything they have to the poor or who dedicated their lives to doing good in other ways. These people are often rather met by lack of understanding, criticism or ridicule than societal approval or reward. Whilst this is anecdotal, it seems that these examples put considerable pressure on Kant's notion that we

Act-Consequentialism is not based on a material principle, if by "material principle" we mean one's own happiness and it seems that this is the dominant meaning in Kant or at least the reason why Kant repudiates these principles so sternly. The existence of Act-Consequentialism thus shows that the distinction between material and formal is not exhaustive. I suspect that in the background of Kant's dichotomy is his view that "[t]o be happy is necessarily the demand of every rational but finite being and therefore an unavoidable determining ground of its faculty of desire" (V:25.12-4). Since we cannot but have our personal happiness as a determining ground, Kant thinks of the pursuit of personal happiness as a merely technical issue (V:25.37-26.33); an issue that requires only instrumental reasoning on the part of the agent. He falsely assumes this as the model for *all* material reasoning. He is unaware that there is a morally very important distinction even within the realm of material reasoning: My own good vs. the good of everyone.

Kant, of course, is aware that once we make a formal principle the supreme principle of morality we can derive material content in the form of obligatory ends from it (see for instance V:34.32-35.5). The obligatory end of beneficence is, however, different from *impartial* happiness as it does not include the agent's *own* happiness (V:37.5-7, VI:386.1-7) and it allows for partiality (see for instance VI:390.12).

Kant does not address Act-Consequentialism as we understand it today or its underlying ideas such as impartial promotion of the good and maximization. Kant's famous point concerning the contingency of consequences is intended as a point against eudaimonism and a proper discussion of impartial promotion of the good is strangely absent from Kant's works. Of course, Kant's point regarding the indeterminacy of consequences as they affect the agent herself can be generalized to a problem for impartial forms of Consequentialism (the Cluelessness Objection). I will come back to this point below when I give some indication of how Kantians should engage the Consequentialist.

III. The Fact of Reason and Gallows 2

either act from formal moral principles or from self-interest. The burden is on Kant to show that despite appearances people who make great personal sacrifices to benefit the worst-off always either act from duty in his sense or self-love. I thank an anonymous referee for urging me to spell this point out more.

After having clarified who Kant considers his main opponent and that he was not directly concerned with Act-Consequentialism, I will now show how these discoveries impact and enrich our understanding of one of the key elements of Kant's ethics: his doctrine of the Fact of Reason. My discussion will serve two purposes. Firstly, it further supports my reading so far and shows how my reading can help us to better understand other aspects of Kant's philosophy. Secondly, as I will argue in the next section, it helps us defend Kant against a prominent brand of criticism, which is based on the notion that the Fact of Reason is supposed to establish or vindicate specifically Kant's ethics. The point I will make in this and the next section is that the Fact of Reason should be understood as vindicating Kant's ethics over eudaimonistic approaches. It does this by casting doubt on psychological eudaimonism, a doctrine on which eudaimonism, as Kant understands it, hinges. Since Kant thinks that all of his opponents are eudaimonists, he believes that the Fact of Reason shows that his theory is to be preferred over *all* other theories. However, since there are non-eudaimonist ethical theories, such as Act-Consequentialism, this is not the case. The Fact of Reason is successful against what Kant saw as his main opponents, but it achieves less than Kant assumes. I will propose that we limit the scope of the Fact of Reason and thus arrive at a defensible account of this conception.

The Fact of Reason has the vital role to "prove" (V:42.4-8) that pure practical reason can be practical on its own, i.e., that rational principles can prescribe actions and provide a sufficient motive for them.[39] Showing this is Kant's main goal in his Second Critique.[40] Kant variously characterizes the Fact of Reason as a consciousness of the moral law (V:31.24), consciousness of "freedom of the will" (V:42.9), the moral law itself (V:47.11-13, 91.19-29), as well as "autonomy in the principle of morality" (V:42.6-8). I shall not discuss whether there are tensions between these different characterizations.[41] In Gallows 2, an example I will shortly discuss in detail, it becomes apparent that the Fact of

39 See also V:3.10-13, 31.3234, 32.1-7, 42.4-8, 91.18-29.

40 See V:3.2-24, 15.16-18, 31.1-37, 45.15-22.

41 Allison (1990, 231-3) plausibly argues that the Fact is *consciousness* of the moral law, since Kant's argumentative strategy would be question begging, if the Fact was identical with the moral law. In what follows, I will bracket many of the controversial interpretative issues that the Fact of Reason raises, such as whether the Fact should be understood as something given or something made.

Reason is best understood as an agent's *consciousness that she can abstract from self-love and nonetheless be motivated to act* (see also V:55.15-19, 92.9-11). This is significantly broader than how Kant sometimes speaks of the Fact as a consciousness specifically of the *moral law* (V:30.34, 55.17, 121.3-4).[42]

It is important that in the case that most clearly illustrates the Fact of Reason, the famous Gallows 2 of the Second Critique's §6, the specific normative principle underlying the judgment that I ought to do something is left unspecified. The example only shows that self-love cannot be the source of this judgment. Immediately preceding this case, Kant argues for the reciprocity between freedom and moral obligation (V:28.31-29.25, see also IV:447.6-7, 452.35-453.2). This reciprocity as such, however, does not establish that we are, in fact, under moral obligation and thus free. Gallows 2 seeks to demonstrate this and thus to confirm by "experience" (V:30.21) the viability of Kant's strategy to start from our consciousness of moral obligation and derive freedom from this consciousness. In the section following Gallows 2, Kant states that the awareness that Gallows 2 elicits or demonstrates "may be called a fact of reason" (V:31.24). Gallows 2 is thus intimately connected to Kant's doctrine of the Fact of Reason.[43]

Here is the case in full:

> But experience also confirms this order of concepts in us. Suppose someone asserts of his lustful inclination that, when the desired object and the opportunity are present, it is quite irresistible to him; ask him whether, if a gallows were erected in front of the house where he finds this opportunity and he would be hanged on it immediately after gratifying his lust, he would not then control his inclination. One need not conjecture very long what he would

42 Allison (2013, 137) and Schönecker (2014, 60) think that the Fact is consciousness of the *Categorical Imperative*, i.e., of a specific ethical principle. Kleingeld (2010, 56) makes a convincing case against this: Kant believes that he is the first person to formulate the Categorical Imperative clearly. Consciousness of this specific principle cannot be part of every agent's understanding of morality. She argues further that the Fact of Reason should be understood as "consciousness of the fundamental law of pure practical reason", not of any specific (moral) principle (KLeingeld 2010, 66).

43 This is also recognized in the literature. Grenberg (2013, 18-9, 56, ch.7) for instance discusses the case at length and puts enormous weight on it as evidence for her phenomenological reading of the Fact of Reason. I should note that the Fact is mentioned for the first time in V:6.12 as a "Factum", though it is not discussed there. A discussion only takes place following §6.

> reply. But ask him whether, if his prince demanded, on pain of the same immediate execution, that he give false testimony against an honourable man whom the prince would like to destroy under a plausible pretext, he would consider it possible to overcome his love of life, however great it may be. He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. He judges, therefore, that he can do something because he is aware that he ought to do it and cognizes freedom within him, which, without the moral law, would have remained unknown to him. (V:30.21-35)

The first case demonstrates how in matters of prudence an agent can easily change his mind when the expected outcome of an action changes (see also V:22.33-24.10). This, however, exhibits merely practical freedom, i.e., the ability to forego immediate impulses and inclinations for the sake of one's own total happiness (see also A/B:801-2/829-30). In the second case, Gallows 2, whether the agent would succumb to the threat of the prince or change his mind and refuse, he "will perhaps not venture to assert". The agent will, however, judge "without hesitation" that it is *possible* to decide against his self-interest. His judgment reveals that he believes that he can do something, because he is aware that he ought to. Rational agents, when confronted with cases that require actions for the sake of morality and against their self-interest, admit that they are compelled by morality in a way radically different from their own happiness. Their judgments reveal awareness of a special kind of normativity, one not based on self-love.

Gallows 2 is supposed to demonstrate or confirm that agents can abstract from self-love and act on a rational principle. Immediately following Gallows 2, Kant formulates the "**Fundamental Law of Pure Practical Reason**" as an "unconditional […] a priori […] categorical practical proposition by which the will is objectively determined" (V:30.37-31.9). As I will argue in the next section, the principle Kant has in mind here, the Categorical Imperative, is only one possible way of spelling out the intuition Gallows 2 reveals. There is a mismatch between the consciousness Gallows 2 does in fact reveal (that we can act on other determining grounds than self-love) and what Kant thinks it does reveal (that we can act on duty understood as the Categorical Imperative).

Gallows 2 as Kant stipulates it leaves consequences underdetermined. In fact, that Kant does not clearly stipulate what the overall consequences would be if the agent gives

in or refuses lends further credence to my claim that Kant is not concerned with Consequentialism.[44] He could have stipulated a case in which the consequences slightly favor lying in court, but in which a rational agent still feels intuitively drawn to refusing the prince, since duty, honesty, a fair justice system, etc. matter beyond consequences. Kant, however, does not do this. It is not clear which course of action would have the better overall consequences in Gallows 2. However, if we make the plausible assumption that refusing the prince's request would have better overall consequences – for instance because lying in court comes with additional costs such as undermining the legal system and weakening political opposition to the prince – the Act-Consequentialist would here give the same verdict as the Kantian. In addition, some Consequentialists might have a conception of the good that includes honesty, personal integrity and resisting tyranny and they might assign great importance to these goods and thus concur with Kant.

There is an obvious Kantian objection to the way I handle Gallows 2. Act-Consequentialism *may* yield the same verdict as the Kantian view – but whether it does is contingent on empirical conditions. Suppose that lying in this case does not undermine the legal system, and that the act of lying has the consequence that both the defamer and the prince enjoy enormous happiness, which far outweighs the unhappiness caused to the defamed and any other parties. Then Act-Consequentialism, unless their conception of goodness extends to honesty, integrity, justice, etc., would say that the right course of action is lying. I will address this in the next section and argue that the Act-Consequentialist's way of issuing ethical directives is still relevantly different from how eudaimonists would think about Gallows 2. Act-Consequentialists' reasoning about Gallows 2 resembles the way Kantians would think about the case insofar as both would apply rational, non-self-interested principles and be careful not to accord one's personal well-being an elevated status.

It cannot be an accident that Kant leaves consequences in Gallows 2 underdetermined. Kant is fully aware that it is the job of the philosopher to set up fictional cases specifically in the light of what she wants to confirm or illustrate. In one of the few passages in which Kant speaks explicitly about how the ethicist should go about doing

44 See also Kant's example of "a man who is otherwise honest" but tempted to lie (V:92.36) and who forsakes his intention to lie once he "recognizes the worthlessness of the liar". No mention is made here of the normative principle responsible for this cognition and what the role of consequences is.

her business, he states that she can "like a chemist at any time set up an experiment with every human being's practical reason" (V:92.28-9). Experiments are artificial and controlled situations explicitly created to isolate the structure under investigation from interfering factors.[45] In the Gallows 2 thought experiment, the situation is carefully crafted such that morality unambiguously requires one option, whereas prudence unambiguously favours another option. Kant here wants to uncover what our consciousness of moral necessitation reveals about our freedom, not what principle best systematizes this moral consciousness.

There is, however, one case in Kant that brings out anti-Consequentialist intuitions. The Deposit Case in the *Common Saying* essay[46]: An agent is entrusted with a deposit and could embezzle it to his own advantage and to the advantage of his family (VIII:286.17-29). Kant stipulates the case such that it would not only be best for the agent, but also be impartially better to embezzle the deposit, since the agent's family is needy whereas the deposit's rightful owners are affluent and "adding anything to their resources would be equivalent to throwing it into the sea" (VIII:286.28-9). Yet, according to Kant, even a boy would declare that keeping the deposit is wrong regardless of any consequences, simply because it is a violation of duty (VIII:286.31-3).

The Deposit Case has a very specific target: It is directed against Christian Garve who is for Kant a paradigm of eudaimonism, since he denies that duty can motivate without any prospect of personal happiness in return (VIII:284.9-285.22). Kant points out that it is not clear that the agent would be successful at "furthering his own *happiness* by giving up the deposit" (VIII:286.33-34). Once more Kant focuses on how a course of action would affect the agent's own happiness, not impartial happiness. Handing over the deposit might result in reward and embezzling it might be an ineffective means to escape hardship (VIII:286.36-287.10). The outcome is highly uncertain, but if an agent "asks himself what his duty is in this matter, he is not at all perplexed about what answer to give but certain on the spot what he has to do" (VIII:287.15-7).

We can assume that what does the work for Kant in Deposit is that the example shows that we acknowledge that certain things are morally forbidden, even though, on a

[45] I elaborate on this method in Sticker (2016b, sec.5 and 2017).

[46] A shorter version is already presented in the Second Critique (V:27.27-28.3). However, here consequences are not even mentioned.

first glance, they appear to be in our *own* interest. Deposit, if convincing, would also constitute an effective means to bring out intuitions against impartial Consequentialism. However, I do not think that Deposit is as intuitive as Gallows 2. I take it that there is considerable intuitive pull to opt for the good over the right here[47], and that examples such as Deposit, alongside the more infamous Murderer at the Door, are responsible for the frequent charge against Kant that his ethics suffers from rigorism and context-insensitivity.[48] This indicates a larger problem for Kant: An intuitively plausible, anti-eudaimonist case might not establish Kantianism over Consequentialism and a case that would establish Kantianism over Consequentialism might not be intuitively plausible. I think it was wise that in the context of the Fact of Reason Kant appealed to a case of the former kind – an intuitively plausible case, albeit one that does not support his ethics over all other ethical theories.

I will now turn to the questions of how Gallows 2 plays out in the context of Kant's discussion of eudaimonism and whether the Fact of Reason is feasible.

IV. The Fact of Reason, Eudaimonism and Consequentialism

Many philosophers and even many Kantians are critical of Kant's idea that his Second Critique hinges on a supposed Fact of Reason. The Fact is maligned as a "footstamping" (Guyer 2007, 462), an "act of desperation" [Verzweiflungstat] (Prauss 1983, 67), "moralistic bluster" (Wood 2008, 135), or "a relapse into […] dogmatic rationalism" (Sussman 2008, 52). The main worry behind these criticisms is that, when he realized that his *Groundwork* deduction failed, Kant "bluntly stipulated what he should have carefully argued for" (Kleingeld 2010, 60). I believe that much of this criticism stems from an underlying assumption that the Fact of Reason is supposed to *establish or justify* Kant's ethics, including his version of a supreme principle of morality, the Categorical

[47] It should be noted that our intuitions might be distorted in Deposit, since it presupposes the practice of inheriting, a practice not obviously morally just in our current world of inherited wealth and poverty. It is not clear to me that the affluent heirs have a very strong moral claim on the deposit. Thus Deposit might not be an ideal case for what Kant wishes to show anyway.

[48] Even defenders of Kant, such as Allison (2011, 24), acknowledge that, "the common view is that duties, even supposedly strict ones […] allow for exceptions under certain circumstances". Since the infamous Murderer at the Door has become a topic of its own in Kant scholarship, I shall bracket it for my purpose.

Imperative, in much the same way as the *Groundwork* deduction was supposed to.[49] This puts an immense burden of proof on the Fact. The Fact would have to be an awareness with a very specific content, namely, one that can only be made sense of in the Kantian framework and that replaces a proper philosophical argument (a deduction) for this framework. According to this reading, if someone accepts that there is a Fact of Reason, she rationally has to accept Kant's ethics (or at least strong reasons to do so). The Fact of Reason will appear as less of a footstamping and less desperate, if its function is limited or if it is just one step of a more comprehensive argument for Kant's specific normative principles and general ethical framework, instead of what Kant offers *in lieu* of an argument. This would, on its own, certainly not make the Fact uncontentious but it would go a long way in addressing the worry that Kant in the Second Critique simply assumes what he failed to prove in the *Groundwork*.[50]

To be clear, I do not claim that this worry is all there is to the above criticisms. There are at least two other reasons why philosophers find the Fact of Reason problematic. Firstly, the Fact supposedly establishes the supreme normative authority of morality over principles of self-love. This resonates, for instance, in the moralistic bluster charge.

---

[49] A common way of understanding the *Groundwork* deduction is as a deduction of the *Categorical Imperative* (see, for instance, Paton 1946, 242, Schönecker 1999), i.e., of a specific normative principle. There are, of course, other candidates of the deduction's object. See Sticker (2014, fn.15) for an overview.

[50] See also Ware (2014, 9) who summarizes the foot-stamping and moralistic bluster objection against the Fact thusly: "it certainly sounds as if [Kant] is asserting our moral consciousness as a brute fact. After all, what else could it mean to say our moral consciousness 'forces itself upon us', if not that we have direct, intuitive insight to the truth of the moral law?" Part of the problem here seems to be the object of the cognition: That it is a cognition, which would vindicate Kant's supreme principle of morality over all other principles. My present defense of the Fact is complementary to Ware's (2014, 14) who argues that "Kant is not treating our moral consciousness dogmatically, i.e., as a brute given. Rather, he is treating it as something actual, a matter of fact". Ware is concerned with the *way* we come to know of the Fact, I am concerned with the Fact's content. In a separate paper (Sticker 2018), I argue that even though the Fact appears as self-evident to rational agents, it is possible for Kant to provide further support for his claim that the Fact exists and that it provides the proper starting point for ethical theory. This can be very important when addressing colleagues, such as Garve, who would deny the existence of a Fact of Reason. That Kant can support his notion of a Fact of Reason with further evidence is an important part of my idea that the Fact is more plausible than many philosophers assume, but I cannot make a case for this in the current paper. The current paper focuses on the Fact's *content*.

Secondly, the major upshot of the Fact of Reason is supposedly transcendental freedom.[51] There are thus three things that philosophers worry the Fact is supposed to show but might not be able to: that the Categorical Imperative or Kant's normative principles are the correct moral principles, that morality is of higher authority than self-love and that we are transcendentally free. In what follows, my discussion focuses on the first of these three worries, but it also impacts the second worry, since the argument I will develop also establishes that, according to the common cognition of duty, morality is of superior authority to self-love. Kant's argument, however, leaves open how to best spell out morality and, based on the notion of the Fact of Reason alone, it is unwarranted to assume that Kant's normative ethics must be the best ethical theory. The third issue is a separate point pertaining to Kant's deduction of freedom. I cannot address this deduction, but below I will say something about autonomy.

There are other readings than mine that aim to save the Fact of Reason by putting less argumentative weight on it. O'Neill (2002) and Łuków (1993), for instance, regard the Fact not as part of the Second Critique's main argument but as a mere encore and description of how morality registers in ordinary moral live. In contrast to them, I believe that the Fact is part of Kant's main argument, but it can only be a *part* of an argument that seeks to establish Kant's ethics as the correct moral theory.

I will now argue that the Fact of Reason is supposed to show that eudaimonism is an inadequate ethical theory, but that it leaves open which alternative theory to adopt. Kant believes that his own theory is the only alternative, but Act-Consequentialism is another option. The Fact of Reason only establishes what is common ground amongst non-eudaimonists. In a second step of his argument, Kant then has to show that of those theories compatible with a Fact of Reason his theory is the best.

There are three different ways philosophers can react to the way Kant describes the agent's response in Gallows 2.[52] The first reaction is to accept Kant's description as

[51] I am grateful to an anonymous referee for bringing these additional problems to my attention.

52 A potential fourth reaction would be *intuitionism*. Intuitionists can accept that cases like Gallows 2 reveal that agents believe that they can act on what they see as the moral option. The difference between Intuitionism and a straightforwardly Kantian account is that the intuitionist would presumably not admit "without hesitation" (V:30.32) that she ought to refuse the prince's request. She would do so after weighing

plausible, but to maintain that the response is not rational. Those who wish to pursue this option typically provide an error-theory that explains why an agent judges against her self-interest and even feels motivated by this judgment. This is presumably the way committed Nietzscheans, Freudians, and evolutionary biologists would respond to Gallows 2. I bracket this reaction, since it raises difficult issues about moral skepticism.[53]

The second reaction is to admit that the response of the agent in Gallows 2 counts against one's theory or is something one's theory cannot accommodate. In this case, philosophers might want to deny that the response in Gallows 2 is psychologically plausible or even possible. Eudaimonists, as Kant conceives of them, cannot accept Gallows 2 as Kant describes it, since they think of all moral deliberation along the lines of Gallows 1: Overcoming immediate desires for greater future happiness. As pointed out in sec.I, eudaimonism typically draws on a form of psychological hedonism or psychological eudaimonism: Every rational human pursuit strives for personal happiness. If eudaimonists admit that this is not actually the case (as suggested by Gallows 2), there would be a serious hole in their conception of agency, a conception which supports their normative theory. Kant, by contrast, can admit, and in fact he does admit, that agents normally pursue their personal happiness (V:93.7-10). He only needs to show that agents *can* sacrifice their own well-being for the sake of morality and that they think they *should*. Gallows 2, if the description of the agent's response is plausible and psychologically realistic, supports Kant's theory over eudaimonism.[54]

It should be noted that it is possible that a eudaimonist calculates that fulfilling the prince's request will undermine any possibility of future personal happiness and that hence the gallows is the better choice for her. This eudaimonist would still differ from the reaction described in Gallows 2 in that her reasoning would be structurally like the one in Gallows 1. She would unambiguously change her mind after her calculation because of

---

prima facie duties against each other. This might not be very different from Consequentialism, which also might not issue the right verdict without hesitation (see blow).

53 See instead Schönecker (2014, sec.2) for brief critical discussion of these approaches in the context of the Fact of Reason.

54 Technically, Gallows 2 is not directed against normative hedonism/eudaimonism directly but *psychological* hedonism/eudaimonism. It therefore constitutes a problem for all theories that make this psychological assumption. Kant thinks that all eudaimonist theories make this assumption. It could also be the case that other theories do (see next footnote).

what is in her own best interest. Refusing the prince's request is what she *will* do since anything else is less conducive to her happiness and her own happiness is all that matters. The same holds for a eudaimonism that promises reward for moral actions in an afterlife. Once more the agent would be reasoning about Gallows 2 along the lines of Gallows 1. If an agent admits that Gallows 2 is a situation that requires personal sacrifice without prospect of reward in an afterlife and yet acknowledges that she should and could refuse the prince's request, then this is all Kant needs to put pressure on a theory that draws on the idea that agents can only ever act for the sake of their own happiness.[55] What is of interest for Kant to distinguish his theory from eudaimonism is not so much what *action* an agent judges she should be committing/omitting in this situation, but the way she arrives at her judgment and what she claims grounds this judgment (rational self-interest or something else).

The third reaction is to claim that the response of the agent in Gallows 2 supports one's theory. This option is open to both Kantians and Consequentialists. It is very important that, as I have argued, Gallows 2 does not spell out the normative principle underlying the agent's judgment and does not say that we need specifically the *Categorical Imperative* for this. Kant says that awareness of the moral law "transfers us, in idea" into the world of understanding, and lets us cognize freedom (V:43.30-7, see also V:30.34, VI:49-50.fn) or that we must obtain insight into our freedom from consciousness of moral obligation. We do not need the moral law as Kant conceives of it for this. Any principle that issues verdicts independently from self-love and that can motivate agents to actions is suitable for this task.

I argued in sec.II that Act-Consequentialism can be understood as an ethical theory that advocates a principle that issues verdicts independently of an agent's self-love. Consequentialism does not assume that morality is authoritative because it is the best way for the agent herself to live a good or flourishing life. Act-Consequentialists can endorse the way Kant describes Gallows 2, as this scenario is compatible with their theory and, in fact, supports their theory over other theories based on psychological hedonism/eudaimonism.[56] Gallwos 2 shows that it is possible (conceivable) that agents

[55] I am grateful to Pärttyli Rinne and James Levine for pushing me on the point that eudaimonists might sacrifice their life in Gallows 2.

[56] Certain Act-Consequentialist positions, in particular classical ones, however, endorse psychological

feel motivated by what they acknowledge as morally right, even on a conception of morality that is not self-love based.

However, there is still at least one significant difference between Kantianism and Act-Consequentialism pertaining to Gallows 2. Kant is confident that agents even before they learn about ethical theory are aware of certain absolute prohibitions, such as not to lie and not to bear false testimony (IV:404.8-10, VI:481.22-482.4). He presumably assumes that the verdict in Gallows 2 is the product or expression of such an awareness. Consequentialists, by contrast, might rather assume that the verdict (even if it coincides with the one the Kantian predicts) is the result of calculating foreseeable consequences. A different process of reasoning leads to the verdict, and it could even be the case that the Consequentialist's verdict itself would differ from the Kantian. Depending on circumstances the Consequentialist might judge that lying is the right option.

However, it is important to bear in mind that Kant does not specify the exact process that leads to the verdict. He believes that the verdict is not an expression of the agent's self-love.[57] Even in cases in which the Consequentialist's verdict differs from Kant's, Kant can acknowledge that a person who reasons according to impartial standards is not a

---

hedonism. Bentham, for instance, was even criticized by his fellow Utilitarian Mill (1985, 13) for his supposedly inadequate hedonistc and egoistic explanatory framework. A form of Consequentialism that endorses psychological hedonism would be as incompatible with how Kant describes Gallows 2 as eudaimonism would. I take it, however, that current forms of Consequentialism usually do not make this assumption. This is largely the case because if one is a psychological hedonist but also demands impartial maximization of the good, one needs to assume that it promotes the good maximally if everyone always pursues their own interests. This seems a dubious claim in a world in which the well-off could make a great difference for some of the worst-off by making relatively minor personal sacrifices. See for instance Singer, Lazari-Radek (2015, ch.11.3) who argue that today Consequentalism is more demanding than the early Utilitarians thought, since nowadays we can usually put our resources to a much better (moral) use than spending them on ourselves. Psychological hedonism/eudaimonism is not an attractive option for contemporary Consequentialists. I am grateful to Matthias Hoesch for discussion of classical forms of Consequentialism and their relation to Kant.

[57] Kant emphasizes that an agent "must admit without hesitation that it would be possible for him" to act on his verdict in Gallows 2 (V:30.21-35). This, however, does not imply that figuring out what the right verdict is must happen immediately, without hesitation and without any calculation. In Sticker (2015), I argue that Kant himself sometimes acknowledges that his claim that moral cognition is always easy and immediate is overly optimistic and that he is open to the possibility that moral cognition can be complex.

psychological eudaimonist. This person would disagree with Kant about important normative issues, but she does not share the eudaimonists' idea that agents only ever pursue their personal happiness.

Of course, it introduces a good deal of additional complexity to Gallows 2, if the impartially best option would be lying in court, since then the impartial verdict coincided with the eudaimonist's verdict. Kant would presumably suspect that a self-identified Consequentialist who recommends lying in court as the best option is in fact engaged in an attempt of "rationalizing" (IV:405.14). She tries to make her own pursuit of happiness look morally and rationally justified even if it is not. A solution might be to confront her with other cases in which promoting the impartial good would be costly to her personally, for instance a modified Gallows 2 in which the prince only threatens her with minor inconveniences. This would enable the Kantian to distinguish truly impartial Consequentialists from eudaimonists.[58]

It would be a fruitful place for the Kantian and the Consequentialist to continue their debate by figuring out how to best conceptualize the process that lets common agents reach their verdict when confronted with Gallows 2 type cases. We should bear in mind that the Consequentialist does not have to assume that the agent in Gallows 2 reaches his verdict via a calculation of consequences. The Consequentialist might well concede to the Kantian that agents usually acknowledge moral constraints on their actions, but once she has established common ground with the Kantian she would, in a second step, present philosophical arguments to the effect that these commonsensical constraints are not absolute and should not constrain pursuit of impartial goods.

Let me address another concern that my interpretation raises. Gallows 2 is explicitly presented in the context of Kant's argument for starting with a consciousness of morality and deriving freedom from this consciousness. This raises the question that if we interpret the rational principle underlying the judgment in Gallows 2 as the principle of Act-Consequentialism, does Gallows 2 still reveal that we are transcendentally free or *autonomous* as opposed to merely practically free as Gallows 1 did? Or in other words: Can the principle of Act-Consequentialism be self-imposed? If not, Kant would insist that

[58] Of course, Kant still has to assume that the agent does not simply lie to him. This, however, is a general methodological problem for all cases that seek to bring out people's moral judgments, intuitions, etc.

Act-Consequentialism is as different from his own theory as any eudaimonist theory is, and that it misses the crucial point that allows ethical theory to encourage actions of moral worth: that the supreme principle of morality is not external but the law of an agent's own reason.

For Kant two kinds of principles are not self-imposed. Those in the service of my *self-love* and those imposed on us by *others*.[59] Act-Consequentialism is neither. I have already argued in sec.II that it is an inadequate conception of Act-Consequentialism to see it as a theory in the service of an agent's self-love. Furthermore, that Act-Consequentialists are at odds with many widely-shared moral convictions makes it unlikely that Consequentialist principles are imposed on us by other agents, institutions or the state. Act-Consesequentialism is too revisionary for this.[60] Act-Consequentialism can be self-imposed and an autonomy based theory if by "autonomy" we mean being able to determine what one's duty is based on a rational principle and independently of one's self-love, external authorities, social norms, etc. as well as being able to act on this determination without a motivational push from self-love.[61] This is, in essence, the Kantian conception of autonomy.

The way Kant describes an agent's response in Gallows 2 is plausible under two conditions. Firstly, and this is my main point, we should not expect that from Gallows 2 it follows that Kantian ethics (and no other ethical theory) is correct or that Kantian ethics

---

[59] See for instance IV:433.7-8, 444.24-34, VI:407.19-408.22. See also Reath (2006, 175): "Autonomy requires the capacity to reason and act independently of inclinations [...] It also requires the capacity for independence from certain kinds of social influence".

60 Differences between Consequentialist theories and widely-shared moral convictions pertaining to spending money on luxuries, our obligations towards the globally poor, our treatment of non-rational animals and the distinction between doing and omitting, etc. are discussed at length in Singer (1993).

[61] Historically, many prominent Consequentialist thinkers such as Hume, Bentham, Mill were Compatibilists. Certain contemporary thinkers also stress that "[t]he utilitarian view is [...] compatible with determinism" (Smart 1973, 46). I take it, however, that it is not Consequentialist normative principles that lead Consequentialists to endorse this view but other considerations such as the findings of natural science or assumptions about the nature of moral responsibility. Those who hold Act-Consequentialism as a normative principle could, in principle, accept a more robust, Kantian, notion of freedom without surrendering their normative principles.

has an edge over all other theories. Secondly, Kant does not mean to describe how every single agent would in fact respond to the scenario. He merely means to show that the response he describes is possible and psychologically plausible. Under these two conditions, it seems that there is indeed a Fact of Reason revealed by Gallows 2. The Fact of Reason does support Kant against what he saw as his main adversaries (eudaimonism), but it does not yet commit those who buy into it to Kantianism.

This does not mean that Kantianism and Consequentialism are equally well justified, that nothing can be said for one over the other or that they must converge on the level of normative ethics. After all, my argument only shows that there is common ground between them. There is still plenty of room for debate, once eudaimonism is rejected. Kantians, whilst they should admit that Act-Consequentialism offers a rational non-eudaimonist principle[62], can criticize Consequentialist theories for their tendency to regard rational agents as mere "instrument[s] of impersonal good-making" (Morgan 2009, 20). Furthermore, Kantians can challenge Consequentialists to tell a satisfying story about how the good receives its normative status as good, if not from a formal moral criterion. Moreover, Act-Consequentialism might offer a rational principle that picks out a good action from an impartial perspective, but Consequentialists would still need to tell a story about *how* this principle can become *motivationally* efficacious.[63] Finally, Kant considers it an essential component of ordinary morality that morality is a matter of practical laws that have strict a priori necessity, universality, and certainty. Ordinary agents must be able to grasp them without bothering about endless exceptions and qualifications. Principles of impartial happiness, by contrast, require much empirical knowledge; they lack certainty and necessity; they have merely comparative universality and thus allow for endless exceptions and qualifications. Consequently, they are difficult

[62] Kant does not deny that there are other *rational* principles than the Categorical Imperative. The principle of perfection is for instance a rational principle (V:41.9-38). This principle is, however, still ultimately in the service of self-love.

63 Timmermann (2007b, 161fn.2) thinks that Consequentialism is to be ruled out as an ethical theory if reason "does not and cannot command us to act for the sake of a law that commands us to produce good consequences". Similar worries are articulated by Scanlon (1982, sec.3). Timmermann, however, as Forschler (2013, 101-2) points out, seems to think that acting because of good consequences would always be acting on inclinations never on an impartial principle.

to grasp for ordinary agents.[64] This, of course, is the Cluelessness Objection I mentioned earlier. It could be the case that Kantians can show that Act-Consequentialism, whilst it advocates a rational principle that is superior to eudaimonism, does not advocate the *best* rational principle or the rational principle that we have most reasons to adopt. However, this must be the result of additional and substantive arguments. It is in no way settled by the Fact of Reason.

Conclusion

The yield of my paper for Kantians is to have clarified what Kant intends the Fact of Reason to establish and what it can plausibly establish. The Fact of Reason is less contentious than many of Kant's critics believe since we should not expect it to establish Kantianism. Instead, the Fact only counts against a specific competing view of morality that was particularly prominent at the time and that Kant saw as his main opponent. It is less contentious and less of a footstamping than many of its critics believe.

The yield of my paper for Consequentialists is more modest. Whilst both Kantians and Consequentialists can (and should) accept the Fact of Reason, we should keep in mind that the Fact does not have to play the same role for both theories. From the Second Critique onwards the Fact is an essential part of Kant's project, whereas Consequentialists are not committed to the notion that the Fact is essential for their theory. However, they can employ it as an argument against forms of egoism that they might want to set themselves apart from.

The yield of my paper for the debate about Kantian Consequentialism is that I have shown that Kant's ethics and Act-Consequentialism share an important abstract feature, namely, that they both establish what is to be done from a rational principle and not based on what is in the (reflected) self-interest of an agent, as Kant thinks eudaimonism does. Act-Consequentialism is in some of its fundamental features closer to Kant's ethics than usually assumed. This is no accident, given that Kant is focused on refuting and setting himself apart from a theory that clashes with impartial Act-Consequentialism as well as with his own. It is an interesting question for further debate how Kant's ethics and the

[64] I thank two anonymous referees for urging me to clarify how much room there still is for critical debates between Kantians and Consequentialists on my account. I think there is still plenty.

arguments Kant himself provides for his own normative principles would look if Kant had taken seriously the possibility that there can be other non-self-love based rational principles than the Categorical Imperative, and that we can envisage goods the promotion of which is not in the self-interest of the agent who is called upon to promote them.

Literature

Kant's Works are quoted volume:page.line of the *Academy Edition*. Translations follow the *Cambridge Edition of the Works of Immanuel Kant,* edited by P. Guyer and A. W. Wood.

Allison, H. 1990. *Kant's Theory of Freedom*. Cambridge.
– 2011. *Kant's* Groundwork of the Metaphysics of Morals. *A Commentary*. Oxford, New York.
– 2013. "Autonomy in Kant and German Idealism". In *Kant on Moral Autonomy*. Ed. O. Sensen. Cambridge, 129-46.

Anscombe, G.E.M. 1958. "Modern Moral Philosophy". *Philosophy* 33, 1–19.

Barnes, J. (ed.) 1984. *The Complete Works of Aristotle*. Princeton.

Benden, M. 1972. *Christian August Crusius. Wille und Verstand als Prinzipien des Handelns*. Bonn.

Bentham, J. 1961. *An Introduction to the Principles of Morals and Legislation*, Garden City.

Crusius, C.A. [3]1767. *Anweisung vernünftig zu leben, Darinen nach Erklärung der Natur des Menschlichen Willens die natürlichen Pflichten und allgemeinen Klugheitslehren im richtigen Zusammenhang vorgetragen werden*. Leipzig.
– 1772. *Kurzer Begriff der Moraltheologie, oder nähere Erklärung der praktischen Lehren des Christentums*. Leipzig.

Cummiskey, D. 1990. "Kantian Consequentialism". *Ethics* 100(3), 586- 615.
– 1996. *Kantian Consequentialism*. New York, Oxford.

Dean, R. 2000. “Cummiskey’s Kantian Consequentialism”. *Utilitas* 12(1), 25-40.

DeWitt, J. 2014. “Respect for the Moral Law: The Emotional Side of Reason”. *Philosophy* 89(1), 31-62.

Forschler, S. 2013. “Kantian and Consequentialist Ethics: The Gap can be Bridged”. *Metaphilosophy* 44(1-2), 88-104.

Fugate, C. 2009. “Moral Individuality and Moral Subjectivity in Leibniz, Crusius, and Kant”. In *Cultivating Personhood: Kant and Asian Philosophy*. Ed. S. Palmquist. Berlin, New York, 273-84.

Gregor, M. J. 1996. *Practical Philosophy (The Cambridge Edition of the Works of Immanuel Kant in Translation)*. Cambridge.

Grenberg, J. 2013. *Kant’s Defense of Common Moral Experience. A Phenomenological Account.* Cambridge.

Guyer, P. 2007. “Naturalistic and Transcendental Moments in Kant’s Moral Philosophy”. *Inquiry* 50(5), 444-64.

Hare, R. 1993. “Could Kant Have Been a Utilitarian?”. *Utilitas* 5, 1-16.

Hill, T. JR. 1999. “Happiness and Human Flourishing in Kant’s Ethics”. *Social Philosophy and Policy* 16(1), 143-75.

Hoesch, M., Sticker, M. 2017. “Parfit über Kantianismus und Konsequentialismus”. In *Worauf es ankommt – Derek Parfits praktische Philosophie*. Ed. M. Hoesch, S. Muders, M. Rüther. Hamburg, 27-62.

Kleingeld, P. 2010. “Moral consciousness and the 'fact of reason'”. In *Kant's* Critique of Practical Reason*. A Critical Guide*. Ed. A Reath, J. Timmermann. Cambridge, 55-72.

Lenman, J. 2000. “Consequentialism and Cluelessness”. *Philosophy and Public Affairs* 29(4), 342-70.

Łuków, P. 1993. “The Fact of Reason: Kant’s Passage to Ordinary Moral Knowledge”. *Kant-Studien* 84, 203–21.

Lyons, D. 1965. *Forms and Limits of Utilitarianism*. Oxford.

MacFarquhar, L. 2015. *Strangers Drowning. Voyages to the Brink of Moral Extremity*. London.

Mill, J.S. 1985. “Remarks on Bentham’s Philosophy”. In *Essays on Ethics, Religion and Society*. Ed. J. Robson. Toronto, London, 3-18.
– 1998. *Utilitarianism*. New York.

Moran, K. 2012. *Community and Progress in Kant's Moral Philosophy*. Washington, DC.

Morgan, S. 2009. “Can there be a Kantian consequentialism?”. *Ratio* 22(1), 19–40.

O’Neill, O. 2002. “Autonomy and the Fact of Reason in the Kritik der praktischen Vernunft”. In *Immanuel Kant: Kritik der praktischen Vernunft*. Ed. O. Höffe. Berlin, 81–97.

Paton, H. J. 1946. *The Categorical Imperative: A study in Kant’s Moral Philosophy*. London.

Parfit, D. 2011. *On What Matters*. Oxford.

Pistorius, H.A. 1786. “Rezension der »Grundlegung zur Metaphysik der Sitten«”. *Allgemeine deutsche Bibliothek* 66(2), 447–63.

Prauss, G. 1983. *Kant über Freiheit als Autonomie*. Frankfurt a.M.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA.

Reath, A. 2006. *Agency and Autonomy in Kant's Moral Theory. Selected Essays*. Oxford.

Schmucker, J. 1961. *Die Ursprünge der Ethik Kants in seinen vorkritischen Schriften und Reflektionen*. Meisenheim, Glan.

Scanlon, T. M. 1982. "Contractualism and Utilitarianism". In *Utilitarianism and Beyond*. Ed. A. Sen, B. Williams. Cambridge, 103–28.

Schönecker, D. 1999. *Kant: Grundlegung III. Die Deduktion des kategorischen Imperativs*. Freiburg, München.
– 2014. "Das gefühlte Faktum der Vernunft. Skizze einer Interpretation und Verteidigung". In *Affektivität und Ethik bei Kant und in der Phänomenologie*. Ed. I. Römer. Berlin, Boston, 55-79.

Sidgwick, H. 1907. *The Methods of Ethics*, Seventh Edition. London.

Singer, P. 1971. "Famine, Affluence, and Morality" *Philosophy and Public Affairs* 1, 229-43.
– 1993. *Practical Ethics*. Cambridge.

Sinnott-Armstrong, W. 2015. "Consequentialism". In *The Stanford Encyclopedia of Philosophy*. Ed. E. Zalta. Stanford.
URL <http://plato.stanford.edu/archives/win2015/entries/consequentialism/>.

Smart, J.J.C. 1973. "An outline of a system of utilitarian ethics". In *Utilitarianism for and against.* Ed. J. Smart, B. Williams. Cambridge.

Sticker, M. 2014. "How can Common Human Reason confirm the Correctness of the Deduction in *Groundwork* III – and Why does it Matter?". *Hegel Bulletin* 35(2), 228-51.
– 2015. "The Moral-Psychology of the Common Agent – A Reply to Ido Geiger". *British Journal for the History of Philosophy* 23(5), 976-89.
– 2016a. "Parfit und Kant über vernünftige Zustimmung". *Zeitschrift für praktische Philosophie* 3(2), 221–54.

– 2016b. “Kant on Engaging Other Agents and Observing Reason at Work”. *History of Philosophy Quarterly* 33(4), 347-73.
– 2017. “Experiments in Ethics? Kant on Chemistry, Means of Education and Methods of Practical Philosophy”. *Idealistic Studies* 46(1), 41-63.
– 2018. “Common Human Reason and the Fact of Reason”. In *Proceedings of the 12. International Kant Congress Nature and Freedom*. Ed. V. Waibel, M. Ruffing. Berlin.
– forthcoming. “Sleeping Conscience – Crusius on Moral Fallibility”. In *Crusius*. Ed. A. Hahmann, F. Grunert. Berlin.

Sussman, D. 2008. “From Deduction to Deed: Kant's Grounding of the Moral Law”. *Kantian Review* 13(1), 52-81.

Timmermann, J. 2005. “Why Kant Could not Have Been a Utilitarian”. *Utilitas* 17(3), 243-64.
– 2007a. “Simplicity and Authority: Reflections on Theory and Practice in Kant’s Moral Philosophy”. *Journal of Moral Philosophy* 4(2), 167-83.
– 2007b. *Kant’s* Groundwork of the Metaphysics of Morals*. A Commentary*. Cambridge.
– 2015. “What’s Wrong With ‘Deontology’?”. *Proceedings of the Aristotelian Society* 115, 75-92.

Ware, O. 2014. “Rethinking Kant’s fact of reason’’. *Philosopher’s Imprint* 14, 1–21.

Wood, A. 2000. “Kant vs. Eudaimonism”. In *Kant’s Legacy: Essays Dedicated to Lewis White Beck*. Ed. P. Cicovacki. Rochester, 261-83.
– 2008. *Kantian Ethics*. Cambridge.